DESCRIPTION

VOICE/MUSICAL SOUND ENCODING DEVICE AND VOICE/MUSICAL SOUND ENCODING METHOD

Technical Field

[0001] The present invention relates to a voice/musical tone coding apparatus and voice/musical tone coding method that perform voice/musical tone signal transmission in a packet communication system typified by Internet communication, a mobile communication system, or the like.

Background Art

[0002] When a voice signal is transmitted in a packet communication system typified by Internet communication, a mobile communication system, or the like, compression and coding technology is used to increase transmission efficiency. To date, many voice coding methods have been developed, and many of the low bit rate voice coding methods developed in recent years have a scheme in which a voice signal is separated into spectrum information and detailed spectrum structure information, and compression and decoding is performed on the separated items.

[0003] Also, with the ongoing development of voice telephony environments on the Internet as typified by IP telephony, there is a growing need for technologies

that efficiently compress and transfer voice signals.

[0004] In particular, various schemes relating to voice coding using human auditory masking characteristics are being studied. Auditory masking is the phenomenon whereby, when there is a strong signal component contained in a particular frequency, an adjacent frequency component cannot be heard, and this characteristic is used to improve quality.

[0005] An example of a technology related to this is the method described in Non-Patent Literature 1 that uses auditory masking characteristics in vector quantization distance calculation.

[0006] The voice coding method using auditory masking characteristics in Patent Literature 1 is a calculation method whereby, when a frequency component of an input signal and a code vector shown by a codebook are both in an auditory masking area, the distance in vector quantization is taken to be 0.

Patent Document 1: Japanese Patent Application Laid-Open No.HEI 8-123490 (p.3, FIG.1)

Disclosure of Invention

Problems to be Solved by the Invention

[0007] However, the conventional method shown in Patent Literature 1 can only be adapted to cases with limited input signals and code vectors, and sound quality performance is inadequate.

[0008] The present invention has been implemented taking into account the problems described above, and it is an object of the present invention to provide a high-quality voice/musical tone coding apparatus and voice/musical tone coding method that select a suitable code vector that minimizes degradation of a signal that has a large auditory effect.

Means for Solving the Problems

[0009] In order to solve the above problems, a voice/musical tone coding apparatus of the present invention has a configuration that includes: a quadrature transformation processing section that converts a voice/musical tone signal from time components to frequency components; an auditory masking characteristic value calculation section that finds an auditory masking characteristic value from the aforementioned voice/musical tone signal; and a vector quantization section that performs vector quantization changing an aforementioned frequency component and the calculation method of the distance between a code vector found from a preset codebook and the aforementioned frequency component based on the aforementioned auditory masking characteristic value.

Advantageous Effect of the Invention

[0010] According to the present invention, by performing

quantization changing the method of calculating the distance between an input signal and code vector based on an auditory masking characteristic value, it is possible to select a suitable code vector that minimizes degradation of a signal that has a large auditory effect, and improve input signal reproducibility and obtain good decoded voice.

Brief Description of Drawings

[0011]

FIG.1 is a block configuration diagram of an overall system that includes a voice/musical tone coding apparatus and voice/musical tone decoding apparatus according to Embodiment 1 of the present invention;

FIG.2 is a block configuration diagram of a voice/musical tone coding apparatus according to Embodiment 1 of the present invention;

FIG.3 is a block configuration diagram of an auditory masking characteristic value calculation section according to Embodiment 1 of the present invention;

FIG.4 is a drawing showing a sample configuration of critical bandwidths according to Embodiment 1 of the present invention;

FIG.5 is a flowchart of a vector quantization section according to Embodiment 1 of the present invention;

FIG.6 is a drawing explaining the relative positional relationship of auditory masking

characteristic values, coding values, and MDCT coefficients according to Embodiment 1 of the present invention;

FIG.7 is a block configuration diagram of a voice/musical tone decoding apparatus according to Embodiment 1 of the present invention;

FIG.8 is a block configuration diagram of a voice/musical tone coding apparatus and voice/musical tone decoding apparatus according to Embodiment 2 of the present invention;

FIG.9 is a schematic configuration diagram of a CELP type voice coding apparatus according to Embodiment 2 of the present invention;

FIG.10 is a schematic configuration diagram of a CELP type voice decoding apparatus according to Embodiment 2 of the present invention;

FIG.11 is a block configuration diagram of an enhancement layer coding section according to Embodiment 2 of the present invention;

FIG.12 is a flowchart of a vector quantization section according to Embodiment 2 of the present invention;

FIG.13 is a drawing explaining the relative positional relationship of auditory masking characteristic values, coded values, and MDCT coefficients according to Embodiment 2 of the present invention;

FIG.14 is a block configuration diagram of a decoding section according to Embodiment 2 of the present invention;

FIG.15 is a block configuration diagram of a voice signal transmitting apparatus and voice signal receiving apparatus according to Embodiment 3 of the present invention;

FIG.16 is a flowchart of a coding section according to Embodiment 1 of the present invention; and

FIG.17 is a flowchart of an auditory masking value calculation section according to Embodiment 1 of the present invention.

Best Mode for Carrying Out the Invention

[0012] Embodiments of the present invention will now be described in detail below with reference to the accompanying drawings.

[0013] (Embodiment 1)

FIG.1 is a block diagram showing the configuration of an overall system that includes a voice/musical tone coding apparatus and voice/musical tone decoding apparatus according to Embodiment 1 of the present invention.

[0014] This system is composed of voice/musical tone coding apparatus 101 that codes an input signal, transmission channel 103, and voice/musical tone decoding apparatus 105 that decodes

[0015] Transmission channel 103 may be a wireless LAN, mobile terminal packet communication, Bluetooth, or suchlike radio communication channel, or may be an ADSL, FTTH, or suchlike cable communication channel.

[0016] Voice/musical tone coding apparatus 101 codes input signal 100, and outputs the result to transmission channel 103 as coded information 102.

[0017] Voice/musical tone decoding apparatus 105 receives coded information 102 via transmission channel 103, performs decoding, and outputs the result as output signal 106.

[0018] The configuration of voice/musical tone coding apparatus 101 will be described using the block diagram in FIG.2. In FIG.2, voice/musical tone coding apparatus 101 is mainly composed of: quadrature transformation processing section 201 that converts input signal 100 from time components to frequency components; auditory masking characteristic value calculation section 203 that calculates an auditory masking characteristic value from input signal 100; shape codebook 204 that shows the correspondence between an index and a normalized code vector; gain codebook 205 that relates to each normalized code vector of shape codebook 204 and shows its gain; and vector quantization section 202 that performs vector quantization of an input signal converted to the aforementioned frequency components using the aforementioned auditory masking characteristic value,

and the aforementioned shape codebook and gain codebook.

[0019] The operation of voice/musical tone coding apparatus 101 will now be described in detail in accordance with the procedure in the flowchart in FIG.16.

[0020] First, input signal sampling processing will be described. Voice/musical tone coding apparatus 101 divides input signal 100 into sections of N samples (where N is a natural number), takes N samples as one frame, and performs coding on a frame-by-frame. Here, input signal 100 subject to coding will be represented as x_n ($n = 0, \Lambda, N-1$), where n indicates that this is the n+1'th of the signal elements comprising the aforementioned divided input signal.

[0021] Input signal x_n 100 is input to quadrature transformation processing section 201 and auditory masking characteristic value calculation section 203.

[0022] Quadrature transformation processing section 201 has internal buffers buf_n ($n = 0, \Lambda, N-1$) for the aforementioned signal elements, and initializes these with 0 as the initial value by means of Equation (1).

[0023]

$$buf_n = 0 \quad (n = 0, \cdots, N-1)$$ [Equation 1]

[0024] Quadrature transformation processing (step S1601) will now be described with regard to the calculation procedure in quadrature transformation processing section 201 and data output to an internal buffer.

[0025] Quadrature transformation processing section 201

performs a modified discrete cosine transform (MDCT) on input signal x_n 100, and finds MDCT coefficient X_k by means of Equation (2).

[0026]

$$X_k = \frac{2}{N}\sum_{n=0}^{2N-1} x'_n \cos\left[\frac{(2n+1+N)(2k+1)\pi}{4N}\right] \quad (k=0,\cdots,N-1) \text{[Equation 2]}$$

[0027] Here, k signifies the index of each sample in one frame. Quadrature transformation processing section 201 finds x_n', which is a vector linking input signal x_n 100 and buffer buf_n, by means of Equation (3).

[0028]

$$x'_n = \begin{cases} buf_n & (n=0,\cdots N-1) \\ x_{n-N} & (n=N,\cdots 2N-1) \end{cases} \text{[Equation 3]}$$

[0029] Quadrature transformation processing section 201 then updates buffer buf_n by means of Equation (4).

[0030]

$$buf_n = x_n \quad (n=0,\cdots N-1) \text{[Equation 4]}$$

[0031] Next, quadrature transformation processing section 201 outputs MDCT coefficient X_k to vector quantization section 202.

[0032] The configuration of auditory masking characteristic value calculation section 203 in FIG.2 will now be described using the block diagram in FIG.3.

[0033] In FIG.3, auditory masking characteristic value calculation section 203 is composed of: Fourier transform section 301 that performs Fourier transform processing of an input signal; power spectrum calculation section

302 that calculates a power spectrum from the aforementioned Fourier transformed input signal; minimum audible threshold value calculation section 304 that calculates a minimum audible threshold value from an input signal; memory buffer 305 that buffers the aforementioned calculated minimum audible threshold value; and auditory masking value calculation section 303 that calculates an auditory masking value from the aforementioned calculated power spectrum and the aforementioned buffered minimum audible threshold value.

[0034] Next, auditory masking characteristic value calculation processing (step S1602) in auditory masking characteristic value calculation section 203 configured as described above will be explained using the flowchart in FIG.17.

[0035] The auditory masking characteristic value calculation method is disclosed in a paper by Mr. J. Johnston et al (J.Johnston, "Estimation of perceptual entropy using noise masking criteria", in Proc. ICASSP-88, May 1988, pp.2524-2527).

[0036] First, the operation of Fourier transform section 301 will be described with regard to Fourier transform processing (step S1701).

[0037] Fourier transform section 301 has input signal x_n 100 as input, and converts this to a frequency domain signal F_k by means of Equation (5). Here, e is the natural logarithm base, and k is the index of each sample in one

frame.

[0038]

$$F_k = \sum_{n=0}^{N-1} x_n e^{-j\frac{2\pi kn}{N}} \quad (k = 0, \cdots, N-1) \text{ [Equation 5]}$$

[0039] Fourier transform section 301 then outputs obtained F_k to power spectrum calculation section 302.

[0040] Next, power spectrum calculation processing (step S1702) will be described.

[0041] Power spectrum calculation section 302 has frequency domain signal F_k output from Fourier transform section 301 as input, and finds power spectrum P_k of F_k by means of Equation (6). Here, k is the index of each sample in one frame.

[0042]

$$P_k = \left(F_k^{\mathrm{Re}}\right)^2 + \left(F_k^{\mathrm{Im}}\right)^2 \quad (k = 0, \cdots, N-1) \text{ [Equation 6]}$$

[0043] In Equation (6), F_k^{Re} is the real part of frequency domain signal F_k, and is found by power spectrum calculation section 302 by means of Equation (7).

[0044]

$$F_k^{\mathrm{Re}} = \sum_{n=0}^{N-1} \left[x_n \cos\left(\frac{2\pi kn}{N} \right) \right] \quad (k = 0, \cdots, N-1) \text{ [Equation 7]}$$

[0045] Also, F_k^{Im} is the imaginary part of frequency domain signal F_k, and is found by power spectrum calculation section 302 by means of Equation (8).

[0046]

$$F_k^{\mathrm{Im}} = -\sum_{n=0}^{N-1} \left[x_n \sin\left(\frac{2\pi kn}{N} \right) \right] \quad (k = 0, \cdots, N-1) \text{ [Equation 8]}$$

[0047] Power spectrum calculation section 302 then outputs obtained power spectrum P_k to auditory masking value calculation section 303.

[0048] Next, minimum audible threshold value calculation processing (step S1703) will be described.

[0049] Minimum audible threshold value calculation section 304 finds minimum audible threshold value ath_k in the first frame only by means of Equation (9).

[0050]

$$ath_k = 3.64(k/1000)^{-0.8} - 6.5e^{-0.6(k/1000-3.3)^2} + 10^{-3}(k/1000)^4 \quad (k = 0, \cdots, N-1)$$

[Equation 9]

[0051] Next, memory buffer storage processing (step S1704) will be described.

[0052] Minimum audible threshold value calculation section 304 outputs minimum audible threshold value ath_k to memory buffer 305. Memory buffer 305 outputs input minimum audible threshold value ath_k to auditory masking value calculation section 303. Minimum audible threshold value ath_k is determined for each frequency component based on human hearing, and a component equal to or smaller than ath_k is not audible.

[0053] Next, the operation of auditory masking value calculation section 303 will be described with regard to auditory masking value calculation processing (step S1705).

[0054] Auditory masking value calculation section 303 has power spectrum P_k output from power spectrum

calculation section 302 as input, and divides power spectrum P_k into m critical bandwidths. Here, a critical bandwidth is a threshold bandwidth for which the amount by which a pure tone of the center frequency is masked does not increase even if band noise is increased. FIG.4 shows a sample critical bandwidth configuration. In FIG.4, m is the total number of critical bandwidths, and power spectrum P_k is divided into m critical bandwidths. Also, i is the critical bandwidth index, and has a value from 0 to m-1. Furthermore, bh_i and bl_i are the minimum frequency index and maximum frequency index of each critical bandwidth I, respectively.

[0055] Next, auditory masking value calculation section 303 has power spectrum P_k output from power spectrum calculation section 302 as input, and finds power spectrum B_i calculated for each critical bandwidth by means of Equation (10).

[0056]

$$B_i = \sum_{k=bl_i}^{bh_i} P_k \quad (i = 0, \cdots, m-1) \text{[Equation 10]}$$

[0057] Auditory masking value calculation section 303 then finds spreading function SF(t) by means of Equation (11).

Spreading function SF(t) is used to calculate, for each frequency component, the effect (simultaneous masking effect) that that frequency component has on adjacent frequencies.

[0058]

$$SF(t)=15.81139+7.5(t+0.474)-17.5\sqrt{1+(t+0.474)^2} \quad (t=0,\cdots,N_t-1)$$ [Equation 11]

[0059] Here, N_t is a constant set beforehand within a range that satisfies the condition in Equation (12).

[0060]

$$0 \le N_t \le m$$ [Equation 12]

[0061] Next, auditory masking value calculation section 303 finds constant C_i using power spectrum B_i and spreading function SF(t) added for each critical bandwidth by means of Equation (13).

[0062]

$$C_i=\begin{cases}\sum_{t=N_T-i}^{N_t} B_i \cdot SF(t) & (i<N_t)\\ \sum_{t=0}^{N_t} B_i \cdot SF(t) & (N_t \le i \le N-N_t)\\ \sum_{t=0}^{N-i} B_i \cdot SF(t) & (i>N-N_t)\end{cases}$$ [Equation 13]

[0063] Auditory masking value calculation section 303 then finds geometric mean μ_i^g by means of Equation (14).

[0064]

$$\mu_i^g=10^{\frac{\log\left(\prod_{k=bh_i}^{bl_i} P_k\right)}{bl_i-bh_i}} \quad (i=0,\cdots,m-1)$$ [Equation 14]

[0065] Auditory masking value calculation section 303 then finds arithmetic mean μ_i^a by means of Equation (15).

[0066]

$$\mu_i^a=\sum_{k=bh_i}^{bl_i} P_k \Big/ (bl_i-bh_i) \quad (i=0,\cdots,m-1)$$ [Equation 15]

[0067] Auditory masking value calculation section 303 then finds SFM_i (Spectral Flatness Measure) by means of Equation (16).

[0068]

$$SFM_i = \mu_i^g / \mu_i^a \quad (i = 0, \cdots, m-1) \text{[Equation 16]}$$

[0069] Auditory masking value calculation section 303 then finds constant α_i by means of Equation (17).

[0070]

$$\alpha_i = \min\left(\frac{10 \cdot \log_{10} SFM_i}{-60}, \quad 1 \right) \quad (i = 0, \cdots, m-1) \text{[Equation 17]}$$

[0071] Auditory masking value calculation section 303 then finds offset value O_i for each critical bandwidth by means of Equation (18).

[0072]

$$O_i = \alpha_i \cdot (14.5 + i) + 5.5 \cdot (1 - \alpha_i) \quad (i = 0, \cdots, m-1) \text{[Equation 18]}$$

[0073] Auditory masking value calculation section 303 then finds auditory masking value T_i for each critical bandwidth by means of Equation (19).

[0074]

$$T_i = \sqrt{10^{\log_{10}(C_i) - (O_i/10)} / (bl_i - bh_i)} \quad (i = 0, \cdots, m-1) \text{[Equation 19]}$$

[0075] Auditory masking value calculation section 303 then finds auditory masking characteristic value M_k from minimum audible threshold value ath_k output from memory buffer 305 by means of Equation (20), and outputs this to vector quantization section 202.

[0076]

$$M_k = \max(ath_k, T_i) \quad (k = bh_i, \cdots, bl_i \ , i = 0, \cdots, m-1) \text{[Equation 20]}$$

[0077] Next, codebook acquisition processing (step S1603) and vector quantization processing (step S1604) in vector quantization section 202 will be described in detail using the process flowchart in FIG.5.

[0078] Using shape codebook 204 and gain codebook 205, vector quantization section 202 performs vector quantization of MDCT coefficient X_k from MDCT coefficient X_k output from quadrature transformation processing section 201 and an auditory masking characteristic value output from auditory masking characteristic value calculation section 203, and outputs obtained coded information 102 to transmission channel 103 in FIG.1.

[0079] The codebooks will now be described.

[0080] Shape codebook 204 is composed of previously created N_j kinds of N-dimensional code vectors $code_k^j$ (j = 0, Λ, N_j-1, k = 0, Λ, N-1), and gain codebook 205 is composed of previously created N_d kinds of gain codes $gain^d$ (j = 0, Λ, N_d-1).

[0081] In step 501, initialization is performed by assigning 0 to code vector index j in shape codebook 204, and a sufficiently large value to minimum error $Dist_{MIN}$.

[0082] In step 502, N-dimensional code vector $code_k^j$ (k = 0, Λ, N-1) is read from shape codebook 204.

[0083] In step 503, MDCT coefficient X_k output from quadrature transformation processing section 201 is input, and gain Gain of code vector $code_k^j$ (k = 0, Λ, N-1) read in shape codebook 204 in step 502 is found by means of

Equation (21).

[0084]

$$Gain = \sum_{k=0}^{N-1} X_k \cdot code_k^j \Big/ \sum_{k=0}^{N-1} {code_k^j}^2 \text{ [Equation 21]}$$

[0085] In step 504, 0 is assigned to calc_count indicating the number of executions of step 505.

[0086] In step 505, auditory masking characteristic value M_k output from auditory masking characteristic value calculation section 203 is input, and temporary gain $temp_k$ ($k = 0, \Lambda, N-1$) is found by means of Equation (22).

[0087]

$$temp_k = \begin{cases} code_k^j & \left(\left|code_k^j \cdot Gain\right| \geq M_k\right) \\ 0 & \left(\left|code_k^j \cdot Gain\right| < M_k\right) \end{cases} \quad (k = 0, \cdots, N-1) \text{ [Equation 22]}$$

[0088] In Equation (22), if k satisfies the condition $|code_k^j \cdot Gain| \geq M_k$, $code_k^j$ is assigned to temporary gain $temp_k$, and if k satisfies the condition $|code_k^j \cdot Gain| < M_k$, 0 is assigned to temporary gain $temp_k$.

[0089] Then, in step 505, gain Gain for an element that is greater than or equal to the auditory masking value is found by means of Equation (23).

[0090]

$$Gain = \sum_{k=0}^{N-1} X_k \cdot temp_k \Big/ \sum_{k=0}^{N-1} {temp_k}^2 \quad (k = 0, \cdots, N-1) \text{ [Equation 23]}$$

[0091] If temporary gain $temp_k$ is 0 for all k's, 0 is assigned to gain Gain. Also, coded value R_k is found from gain Gain and $code_k^j$ by means of Equation (24).

[0092]

$$R_k = Gain \cdot code_k^j \quad (k = 0, \cdots, N-1)$$ [Equation 24]

[0093] In step 506, calc_count is incremented by 1.

[0094] In step 507, calc_count and a predetermined non-negative integer N_c are compared, and the process flow returns to step 505 if calc_count is a smaller value than N_c, or proceeds to step 508 if calc_count is greater than or equal to N_c. By repeatedly finding gain Gain in this way, gain Gain can be converged to a suitable value.

[0095] In step 508, 0 is assigned to cumulative error Dist, and 0 is also assigned to sample index k.

[0096] Next, in steps 509, 511, 512, and 514, case determination is performed for the relative positional relationship between auditory masking characteristic value M_k, coded value R_k, and MDCT coefficient X_k, and distance calculation is performed in step 510, 513, 515, or 516 according to the case determination result.

[0097] This case determination according to the relative positional relationship is shown in FIG.6. In FIG.6, a white circle symbol (○) signifies an input signal MDCT coefficient X_k, and a black circle symbol (●) signifies a coded value R_k. The items shown in FIG.6 show the special characteristics of the present invention, and the area from the auditory masking characteristic value found by auditory masking characteristic value calculation section 203 $+M_k$ to 0 to $-M_k$ is referred to as the auditory masking area, and high-quality results closer in terms of the sense of hearing can be obtained changing the

distance calculation method when input signal MDCT coefficient X_k or coded value R_k is present in this auditory masking area.

[0098] The distance calculation method in vector quantization according to the present invention will now be described. When neither input signal MDCT coefficient X_k (○) nor coded value R_k (●) is present in the auditory masking area, and input signal MDCT coefficient X_k and coded value R_k are the same codes, as shown in "Case 1" in FIG.6, distance D_{11} between input signal MDCT coefficient X_k (○) and coded value R_k (●) is simply calculated. When one of input signal MDCT coefficient X_k (○) and coded value R_k (●) is present in the auditory masking area, as shown in "Case 3" and "Case 4" in FIG.6, the position within the auditory masking area is corrected to an M_k value (or in some cases a $-M_k$ value) and D_{31} or D_{41} is calculated. When input signal MDCT coefficient X_k (○) and coded value R_k (●) straddle the auditory masking area, as shown in "Case 2" in FIG.6, the inter-auditory-masking-area distance is calculated as $\beta \cdot D_{23}$ (where β is an arbitrary coefficient). When input signal MDCT coefficient X_k (○) and coded value R_k (●) are both present within the auditory masking area, as shown in "Case 5" in FIG.6, distance D_{51} is calculated as 0.

[0099] Next, processing in step 509 through step 517 for each of the cases will be described.

[0100] In step 509, whether or not the relative

positional relationship between auditory masking characteristic value M_k, coded value R_k, and MDCT coefficient X_k corresponds to "Case 1" in FIG.6 is determined by means of the conditional expression in Equation (25).

[0101]

$$\left(|X_k| \geq M_k\right) \quad and \quad \left(|R_k| \geq M_k\right) \quad and \quad \left(X_k \cdot R_k \geq 0\right) \text{[Equation 25]}$$

[0102] Equation (25) signifies a case in which the absolute value of MDCT coefficient X_k and the absolute value of coded value R_k are both greater than or equal to auditory masking characteristic value M_k, and MDCT coefficient X_k and coded value R_k are the same codes. If auditory masking characteristic value M_k, MDCT coefficient X_k, and coded value R_k satisfy the conditional expression in Equation (25), the process flow proceeds to step 510, and if they do not satisfy the conditional expression in Equation (25), the process flow proceeds to step 511.

[0103] In step 510, error $Dist_1$ between coded value R_k and MDCT coefficient X_k is found by means of Equation (26), error $Dist_1$ is added to cumulative error Dist, and the process flow proceeds to step 517.

[0104]

$$\begin{aligned} Dist_1 &= D_{11} \\ &= |X_k - R_k| \end{aligned} \text{[Equation 26]}$$

[0105] In step 511, whether or not the relative positional relationship between auditory masking

characteristic value M_k, coded value R_k, and MDCT coefficient X_k corresponds to "Case 5" in FIG.6 is determined by means of the conditional expression in Equation (27).

[0106]

$$(|X_k| \geq M_k) \;\; and \;\; (|R_k| \geq M_k) \;\; and \;\; (X_k \cdot R_k < 0)$$ [Equation 27]

[0107] Equation (27) signifies a case in which the absolute value of MDCT coefficient X_k and the absolute value of coded value R_k are both less than or equal to auditory masking characteristic value M_k. If auditory masking characteristic value M_k, MDCT coefficient X_k, and coded value R_k satisfy the conditional expression in Equation (27), the error between coded value R_k and MDCT coefficient X_k is taken to be 0, nothing is added to cumulative error Dist, and the process flow proceeds to step 517, whereas if they do not satisfy the conditional expression in Equation (27), the process flow proceeds to step 512.

[0108] In step 512, whether or not the relative positional relationship between auditory masking characteristic value M_k, coded value R_k, and MDCT coefficient X_k corresponds to "Case 2" in FIG.6 is determined by means of the conditional expression in Equation (28).

[0109]

$$Dist_2 = D_{21} + D_{22} + \beta * D_{23}$$ [Equation 28]

[0110] Equation (28) signifies a case in which the

absolute value of MDCT coefficient X_k and the absolute value of coded value R_k are both greater than or equal to auditory masking characteristic value M_k, and MDCT coefficient X_k and coded value R_k are different codes. If auditory masking characteristic value M_k, MDCT coefficient X_k, and coded value R_k satisfy the conditional expression in Equation (28), the process flow proceeds to step 513, and if they do not satisfy the conditional expression in Equation (28), the process flow proceeds to step 514.

[0111] In step 513, error $Dist_2$ between coded value R_k and MDCT coefficient X_k is found by means of Equation (29), error $Dist_2$ is added to cumulative error Dist, and the process flow proceeds to step 517.

[0112]

$$D_{21} = |X_k| - M_k \text{ [Equation 29]}$$

[0113] Here, β is value set as appropriate according to MDCT coefficient X_k, coded value R_k, and auditory masking characteristic value M_k. A value of 1 or less is suitable for β, and a numeric value found experimentally by subject evaluation may be used. D_{21}, D_{22}, and D_{23} are found by means of Equation (30), Equation (31), and Equation (32), respectively.

[0114]

$$D_{22} = |R_k| - M_k \text{ [Equation 30]}$$

[0115]

$$D_{23} = M_k \cdot 2 \text{ [Equation 31]}$$

[0116]

$$(|X_k| \geq M_k) \quad and \quad (|R_k| < M_k) \text{ [Equation 32]}$$

[0117] In step 514, whether or not the relative positional relationship between auditory masking characteristic value M_k, coded value R_k, and MDCT coefficient X_k corresponds to "Case 3" in FIG.6 is determined by means of the conditional expression in Equation (33).

[0118]

$$\begin{aligned} Dist_3 &= D_{31} \\ &= |X_k| - M_k \end{aligned} \text{ [Equation 33]}$$

[0119] Equation (33) signifies a case in which the absolute value of MDCT coefficient X_k is greater than or equal to auditory masking characteristic value M_k, and coded value R_k is less than auditory masking characteristic value M_k. If auditory masking characteristic value M_k, MDCT coefficient X_k, and coded value R_k satisfy the conditional expression in Equation (33), the process flow proceeds to step 515, and if they do not satisfy the conditional expression in Equation (33), the process flow proceeds to step 516.

[0120] In step 515, error $Dist_3$ between coded value R_k and MDCT coefficient X_k is found by means of Equation (34), error $Dist_3$ is added to cumulative error Dist, and the process flow proceeds to step 517.

[0121]

$$(|X_k| < M_k) \quad and \quad (|R_k| \geq M_k) \text{ [Equation 34]}$$

[0122] In step 516, the relative positional relationship between auditory masking characteristic value M_k, coded value R_k, and MDCT coefficient X_k corresponds to "Case 4" in FIG.6, and the conditional expression in Equation (35) is satisfied.

[0123]

$$(|X_k| < M_k) \quad and \quad (|R_k| < M_k) \text{ [Equation 35]}$$

[0124] Equation (35) signifies a case in which the absolute value of MDCT coefficient X_k is less than auditory masking characteristic value M_k, and coded value R_k is greater than or equal to auditory masking characteristic value M_k. In step 516, error $Dist_4$ between coded value R_k and MDCT coefficient X_k is found by means of Equation (36), error $Dist_4$ is added to cumulative error Dist, and the process flow proceeds to step 517.

[0125]

$$\begin{aligned} Dist_4 &= D_{41} \\ &= |R_k| - M_k \end{aligned} \text{ [Equation 36]}$$

[0126] In step 517, k is incremented by 1.

[0127] In step 518, N and k are compared, and if k is a smaller value than N, the process flow returns to step 509. If k has the same value as N, the process flow proceeds to step 519.

[0128] In step 519, cumulative error Dist and minimum error $Dist_{MIN}$ are compared, and if cumulative error Dist is a smaller value than minimum error $Dist_{MIN}$, the process flow proceeds to step 520, whereas if cumulative error

Dist is greater than or equal to minimum error $Dist_{MIN}$, the process flow proceeds to step 521.

[0129] In step 520, cumulative error Dist is assigned to minimum error $Dist_{MIN}$, j is assigned to $code\_index_{MIN}$, and gain Gain is assigned to error minimum gain $Dist_{MIN}$, and the process flow proceeds to step 521.

[0130] In step 521, j is incremented by 1.

[0131] In step 522, total number of vectors N_j and j are compared, and if j is a smaller value than N_j, the process flow returns to step 502. If j is greater than or equal to N_j, the process flow proceeds to step 523.

[0132] In step 523, N_d kinds of gain code $gain^d$ (d = 0, Λ, N_d-1) are read from gain codebook 205, and quantization gain error $gainerr^d$ (d = 0, Λ, N_d-1) is found by means of Equation (37) for all d's.

[0133]

$$gainerr^d = \left|Gain_{MIN} - gain^d\right| \quad (d = 0, \cdots, N_d - 1) \text{ [Equation 37]}$$

[0134] Then, in step 523, d for which quantization gain error $gainerr^d$ (d = 0, Λ, N_d-1) is a minimum is found, and the found d is assigned to $gain\_index_{MIN}$.

[0135] In step 524, $code\_index_{MIN}$ that is the code vector index for which cumulative error Dist is a minimum, and $gain\_index_{MIN}$ found in step 523, are output to transmission channel 103 in FIG.1 as coded information 102, and processing is terminated.

[0136] This completes the description of coding section 101 processing.

[0137] Next, voice/musical tone decoding apparatus 105 in FIG.1 will be described using the detailed block diagram in FIG.7.

[0138] Shape codebook 204 and gain codebook 205 are the same as those shown in FIG.2.

[0139] Vector decoding section 701 has coded information 102 transmitted via transmission channel 103 as input, and using $code\_index_{MIN}$ and $gain\_index_{MIN}$ as the coded information, reads code vector $codek^{code\_indexMIN}$ ($k = 0, \Lambda, N-1$) from shape codebook 204, and also reads gain code $gain^{gain\_indexMIN}$ from gain codebook 205. Then vector decoding section 701 multiplies $gain^{gain\_indexMIN}$ by $codek^{code\_indexMIN}$ ($k = 0, \Lambda, N-1$), and outputs $gain^{gain\_indexMIN} \times codek^{code\_indexMIN}$ ($k = 0, \Lambda, N-1$) obtained as a result of the multiplication to quadrature transformation processing section 702 as a decoded MDCT coefficient.

[0140] Quadrature transformation processing section 702 has an internal buffer buf_k', and initializes this buffer in accordance with Equation (38).

[0141]

$$buf_k' = 0 \quad (k = 0, \cdots, N-1) \text{[Equation 38]}$$

[0142] Next, decoded MDCT coefficient $gain^{gain\_indexMIN} \times codek^{code\_indexMIN}$ ($k = 0, \Lambda, N-1$) output from MDCT coefficient decoding section 701 is input, and decoded signal Y_n is found by means of Equation (39).

[0143]

$$y_n = \frac{2}{N}\sum_{k=0}^{2N-1} X'_k \cos\left[\frac{(2n+1+N)(2k+1)\pi}{4N}\right] \quad (n=0,\cdots,N-1)$$ [Equation 39]

[0144] Here, X_k' is a vector linking decoded MDCT coefficient $gain^{gain\_indexMIN} \times codek^{code\_indexMIN}$ (k = 0, Λ, N-1) and buffer buf_k', and is found by means of Equation (40).

[0145]

$$X'_k = \begin{cases} buf'_k & (k=0,\cdots N-1) \\ gain^{gain\_index_{MIN}} \cdot code_{k-N}^{code\_index_{MIN}} & (k=N,\cdots 2N-1) \end{cases}$$ [Equation 40]

[0146] Buffer buf_k' is then updated by means of Equation (41).

[0147]

$$buf'_k = gain^{gain\_index_{MIN}} \cdot code_k^{code\_index_{MIN}} \quad (k=0,\cdots N-1)$$ [Equation 41]

[0148] Decoded signal Y_n is then output as output signal 106.

[0149] By thus providing a quadrature transformation processing section that finds an input signal MDCT coefficient, an auditory masking characteristic value calculation section that finds an auditory masking characteristic value, and a vector quantization section that performs vector quantization using an auditory masking characteristic value, and performing vector quantization distance calculation according to the relative positional relationship between an auditory masking characteristic value, MDCT coefficient, and quantized MDCT coefficient, it is possible to select a suitable code vector that minimizes degradation of a

signal that has a large auditory effect, and to obtain a high-quality output signal.

[0150] It is also possible to perform quantization in vector quantization section 202 by applying acoustic weighting filters for the distance calculations in above-described Case 1 through Case 5.

[0151] Also, in this embodiment, a case has been described in which MDCT coefficient coding is performed, but the present invention can also be applied, and the same kind of actions and effects can be obtained, in a case in which post-transformation signal (frequency parameter) coding is performed using Fourier transform, discrete cosine transform (DCT), or quadrature mirror filter (QMF) or suchlike quadrature transformation.

[0152] Furthermore, in this embodiment, a case has been described in which coding is performed by means of vector quantization, but there are no restrictions on the coding method in the present invention, and , for example, coding may also be performed by means of divided vector quantization or multi-stage vector quantization.

[0153] It is also possible for voice/musical tone coding apparatus 101 to have the procedure shown in the flowchart in FIG.16 executed by a computer by means of a program.

[0154] As described above, by calculating an auditory masking characteristic value from an input signal, considering all relative positional relationships of MDCT coefficient, coded value, and auditory masking

characteristic value, and applying a distance calculation method suited to human hearing, it is possible to select a suitable code vector that minimizees degradation of a signal that has a large auditory effect, and to obtain good decoded voice even when an input signal is decoded at a low bit rate.

[0155] In Patent Literature 1, only "Case 5" in FIG.6 is disclosed, but with the present invention, in addition to this, by employing a distance calculation method that takes an auditory masking characteristic value into consideration for all combinations of relationships as shown in "Case 2," "Case 3," and "Case 4," considering all relative positional relationships of input signal MDCT coefficient, coded value, and auditory masking characteristic value, and applying a distance calculation method suited to hearing, it is possible to obtain higher-quality coded voice even when an input signal is quantized at a low bit rate.

[0156] Also, the present invention is based on the fact that actual audibility differs if distance calculation is performed without change and vector quantization is then performed when an input signal MDCT coefficient or coded value is present within the auditory masking area, and when present on either side of the auditory masking area, and therefore more natural audibility can be provided changing the distance calculation method when performing vector quantization.

[0157] (Embodiment 2)

In Embodiment 2 of the present invention, an example is described in which vector quantization using the auditory masking characteristic values described in Embodiment 1 is applied to scalable coding.

[0158] In this embodiment, a case is described below in which, in a two-layer voice coding and decoding method composed of a base layer and enhancement layer, vector quantization is performed using auditory masking characteristic value in the enhancement layer.

[0159] A scalable voice coding method is a method whereby a voice signal is split into a plurality of layers based on frequency characteristics and coding is performed. Specifically, signals of each layer are calculated using a residual signal representing the difference between a lower layer input signal and a lower layer output signal. On the decoding side, the signals of these layers are added and a voice signal is decoded. This technique enables sound quality to be controlled flexibly, and also makes noise-tolerant voice signal transfer possible.

[0160] In this embodiment, a case in which the base layer performs CELP type voice coding and decoding will be described as an example.

[0161] FIG.8 is a block diagram showing the configuration of a coding apparatus and decoding apparatus that use an MDCT coefficient vector quantization method according to Embodiment 2 of the present invention. In FIG.8, the

coding apparatus is composed of base layer coding section 801, base layer decoding section 803, and enhancement layer coding section 805, and the decoding apparatus is composed of base layer decoding section 808, enhancement layer decoding section 810, and adding section 812.

[0162] Base layer coding section 801 codes an input signal 800 using a CELP type voice coding method, calculates base layer coded information 802, and outputs this to base layer decoding section 803, and to base layer decoding section 808 via transmission channel 807.

[0163] Base layer decoding section 803 decodes base layer coded information 802 using a CELP type voice decoding method, calculates base layer decoded signal 804, and outputs this to enhancement layer coding section 805.

[0164] Enhancement layer coding section 805 has base layer decoded signal 804 output by base layer decoding section 803, and input signal 800, as input, codes the residual signal of input signal 800 and base layer decoded signal 804 by means of vector quantization using an auditory masking characteristic value, and outputs enhancement layer coded information 806 found by means of quantization to enhancement layer decoding section 810 via transmission channel 807. Details of enhancement layer coding section 805 will be given later herein.

[0165] Base layer decoding section 808 decodes base layer coded information 802 using a CELP type voice decoding method, and outputs a base layer decoded signal 809 found

by decoding to adding section 812.

[0166] Enhancement layer decoding section 810 decodes enhancement layer coded information 806, and outputs enhancement layer decoded signal 811 found by decoding to adding section 812.

[0167] Adding section 812 adds together base layer decoded signal 809 output from base layer decoding section 808 and enhancement layer decoded signal 811 output from enhancement layer decoding section 810, and outputs the voice/musical tone signal that is the addition result as output signal 813.

[0168] Next, base layer coding section 801 will be described using the block diagram in FIG.9.

[0169] Input signal 800 of base layer coding section 801 is input to a preprocessing section 901. Preprocessing section 901 performs high pass filter processing that removes a DC component, and waveform shaping processing and pre-emphasis processing aiming at performance improvement of subsequent coding processing, and outputs the signal (Xin) that has undergone this processing to LPC analysis section 902 and adding section 905.

[0170] LPC analysis section 902 performs linear prediction analysis using Xin, and outputs the analysis result (linear prediction coefficient) to LPC quantization section 903. LPC quantization section 903 performs quantization processing of the linear prediction coefficient (LPC) output from LPC analysis section 902,

outputs the quantized LPC to combining filter 904, and also outputs a code (L) indicating the quantized LPC to multiplexing section 914.

[0171] Using a filter coefficient based on the quantized LPC, combining filter 904 generates a composite signal by performing filter combining on a drive sound source output from an adding section 911 described later herein, and outputs the composite signal to adding section 905.

[0172] Adding section 905 calculates an error signal by inverting the polarity of the composite signal and adding it to Xin, and outputs the error signal to acoustic weighting section 912.

[0173] Adaptive sound source codebook 906 stores a drive sound source output by adding section 911 in a buffer, extracts one frame's worth of samples from a past drive sound source specified by a signal output from parameter determination section 913 as an adaptive sound source vector, and outputs this to multiplication section 909.

[0174] Quantization gain generation section 907 outputs quantization adaptive sound source gain specified by a signal output from parameter determination section 913 and quantization fixed sound source gain to multiplication section 909 and a multiplication section 910, respectively.

[0175] Fixed sound source codebook 908 multiplies a pulse sound source vector having a form specified by a signal output from parameter determination section 913 by a

spreading vector, and outputs the obtained fixed sound source vector to multiplication section 910.

[0176] Multiplication section 909 multiplies quantization adaptive sound source gain output from quantization gain generation section 907 by the adaptive sound source vector output from adaptive sound source codebook 906, and outputs the result to adding section 911. Multiplication section 910 multiplies the quantization fixed sound source gain output from quantization gain generation section 907 by the fixed sound source vector output from fixed sound source codebook 908, and outputs the result to adding section 911.

[0177] Adding section 911 has as input the post-gain-multiplication adaptive sound source vector and fixed sound source vector from multiplication section 909 and multiplication section 910 respectively, and outputs the drive sound source that is the addition result to combining filter 904 and adaptive sound source codebook 906. The drive sound source input to adaptive sound source codebook 906 is stored in a buffer.

[0178] Acoustic weighting section 912 performs acoustic weighting on the error signal output from adding section 905, and outputs the result to parameter determination section 913 as coding distortion.

[0179] Parameter determination section 913 selects from adaptive sound source codebook 906, fixed sound source

codebook 908, and quantization gain generation section 907, the adaptive sound source vector, fixed sound source vector, and quantization gain that minimize coding distortion output from acoustic weighting section 912, and outputs an adaptive sound source vector code (A), sound source gain code (G), and fixed sound source vector code (F) indicating the selection results to multiplexing section 914.

[0180] Multiplexing section 914 has a code (L) indicating quantized LPC as input from LPC quantization section 903, and code (A) indicating an adaptive sound source vector, code (F) indicating a fixed sound source vector, and code (G) indicating quantization gain as input from parameter determination section 913, multiplexes this information, and outputs the result as base layer coded information 802.

[0181] Base layer decoding section 803 (808) will now be described using FIG.10.

[0182] In FIG.10, base layer coded information 802 input to base layer decoding section 803 (808) is separated into individual codes (L, A, G, F) by demultiplexing section 1001. Separated LPC code (L) is output to LPC decoding section 1002, separated adaptive sound source vector code (A) is output to adaptive sound source codebook 1005, separated sound source gain code (G) is output to quantization gain generation section 1006, and separated fixed sound source vector code (F) is output to fixed

sound source codebook 1007.

[0183] LPC decoding section 1002 decodes a quantized LPC from code (L) output from demultiplexing section 1001, and outputs the result to combining filter 1003.

[0184] Adaptive sound source codebook 1005 extracts one frame's worth of samples from a past drive sound source designated by code (A) output from demultiplexing section 1001 as an adaptive sound source vector, and outputs this to multiplication section 1008.

[0185] Quantization gain generation section 1106 decodes quantization adaptive sound source gain and quantization fixed sound source gain designated by sound source gain code (G) output from demultiplexing section 1001, and outputs this to multiplication section 1008 and multiplication section 1009.

[0186] Fixed sound source codebook 1007 generates a fixed sound source vector designated by code (F) output from demultiplexing section 1001, and outputs this to multiplication section 1009.

[0187] Multiplication section 1008 multiplies the adaptive sound source vector by the quantization adaptive sound source gain, and outputs the result to adding section 1010. Multiplication section 1009 multiplies the fixed sound source vector by the quantization fixed sound source gain, and outputs the result to adding section 1010.

[0188] Adding section 1010 performs addition of the post-gain-multiplication adaptive sound source vector

and fixed sound source vector output from multiplication section 1008 and multiplication section 1009, generates a drive sound source, and outputs this to combining filter 1003 and adaptive sound source codebook 1005.

[0189] Using the filter coefficient decoded by LPC decoding section 1002, combining filter 1003 performs filter combining of the drive sound source output from adding section 1010, and outputs the combined signal to postprocessing section 1004.

[0190] Postprocessing section 1004 executes, on the signal output from combining filter 1003, processing that improves the subjective voice sound quality such as formant emphasis and pitch emphasis, processing that improves the subjective sound quality of stationary noise, and so forth, and outputs the resulting signal as base layer decoded signal 804 (810).

[0191] Enhancement layer coding section 805 will now be described using FIG.11.

[0192] Enhancement layer coding section 805 in FIG.11 is similar to that shown in FIG.2, except that differential signal 1102 of base layer decoded signal 804 and input signal 800 is input to quadrature transformation processing section 1103, and auditory masking characteristic value calculation section 203 is assigned the same code as in FIG.2 and is not described here.

[0193] As with coding section 101 of Embodiment 1, enhancement layer coding section 805 divides input signal

800 into sections of N samples (where N is a natural number), takes N samples as one frame, and performs coding on a frame-by-frame basis. Here, input signal 800 subject to coding will be designated x_n (n = 0, Λ, N-1).

[0194] Input signal x_n 800 is input to auditory masking characteristic value calculation section 203 and adding section 1101. Also, base layer decoded signal 804 output from base layer decoding section 803 is input to adding section 1101 and quadrature transformation processing section 1103.

[0195] Adding section 1101 finds residual signal 1102 $xresid_n$ (n = 0, Λ, N-1) by means of Equation (42), and outputs residual signal 1102 $xresid_n$ to quadrature transformation processing section 1103.

[0196]

$$xresid_n = x_n - xbase_n \quad (n = 0, \cdots, N-1) \text{[Equation 42]}$$

[0197] Here, $xbase_n$ (n = 0, Λ, N-1) is base layer decoded signal 804. Next, the process performed by quadrature transformation processing section 1103 will be described.

[0198] Quadrature transformation processing section 1103 has internal buffers $bufbase_n$ (n = 0, Λ, N-1) used in base layer decoded signal $xbase_n$ 804 processing, and $bufresid_n$ (n = 0, Λ, N-1) used in residual signal $xresid_n$ 1102 processing, and initializes these buffers by means of Equation (43) and Equation (44) respectively.

[0199]

$$bufbase_n = 0 \quad (n = 0, \cdots, N-1) \text{[Equation 43]}$$

[0200]

$$bufresid_n = 0 \quad (n = 0, \cdots, N-1)$$ [Equation 44]

[0201] Quadrature transformation processing section 1103 then finds base layer quadrature transformation coefficient $xbase_k$ 1104 and residual quadrature transformation coefficient $xresid_k$ 1105 by performing a modified discrete cosine transform (MDCT) on base layer decoded signal $xbase_n$ 804 and residual signal $xresid_n$ 1102, respectively. Base layer quadrature transformation coefficient $xbase_k$ 1104 here is found by means of Equation (45).

[0202]

$$Xbase_k = \frac{2}{N}\sum_{n=0}^{2N-1} xbase'_n \cos\left[\frac{(2n+1+N)(2k+1)\pi}{4N}\right] \quad (k = 0, \cdots, N-1)$$

[Equation 45]

[0203] Here, $xbase_n'$ is a vector linking base layer decoded signal $xbase_n$ 804 and buffer $bufbase_n$, and quadrature transformation processing section 1103 finds $xbase_n'$ by means of Equation (46). Also, k is the index of each sample in one frame.

[0204]

$$xbase'_n = \begin{cases} bufbase_n & (n = 0, \cdots N-1) \\ xbase_{n-N} & (n = N, \cdots 2N-1) \end{cases}$$ [Equation 46]

[0205] Next, quadrature transformation processing section 1103 updates buffer $bufbase_n$ by means of Equation (47).

[0206]

$$bufbase_n = xbase_n \quad (n = 0, \cdots N-1)$$ [Equation 47]

[0207] Also, quadrature transformation processing section 1103 finds residual quadrature transformation coefficient $xresid_k$ 1105 by means of Equation (48).

[0208]

$$Xresid_k = \frac{2}{N}\sum_{n=0}^{2N-1} xresid'_n \cos\left[\frac{(2n+1+N)(2k+1)\pi}{4N}\right] \quad (k = 0, \cdots, N-1)$$

[Equation 48]

[0209] Here, $xresid_n'$ is a vector linking residual signal $xresid_n$ 1102 and buffer $bufresid_n$, and quadrature transformation processing section 1103 finds $xresid_n'$ by means of Equation (49). Also, k is the index of each sample in one frame.

[0210]

$$xresid'_n = \begin{cases} bufresid_n & (n = 0, \cdots N-1) \\ xresid_{n-N} & (n = N, \cdots 2N-1) \end{cases}$$ [Equation 49]

[0211] Next, quadrature transformation processing section 1103 updates buffer $bufresid_n$ by means of Equation (50).

[0212]

$$bufresid_n = xresid_n \quad (n = 0, \cdots N-1)$$ [Equation 50]

[0213] Quadrature transformation processing section 1103 then outputs base layer quadrature transformation coefficient $Xbase_k$ 1104 and residual quadrature transformation coefficient $Xresid_k$ 1105 to vector quantization section 1106.

[0214] Vector quantization section 1106 has, as input, base layer quadrature transformation coefficient $Xbase_k$ 1104 and residual quadrature transformation coefficient $Xresid_k$ 1105 from quadrature transformation processing section 1103, and auditory masking characteristic value M_k 1107 from auditory masking characteristic value calculation section 203, and using shape codebook 1108 and gain codebook 1109, performs coding of residual quadrature transformation coefficient $Xresid_k$ 1105 by means of vector quantization using the auditory masking characteristic value, and outputs enhancement layer coded information 806 obtained by coding.

[0215] Here, shape codebook 1108 is composed of previously created N_e kinds of N-dimensional code vectors $coderesid_k^e$ ($e = 0, \Lambda, N_e-1$, $k = 0, \Lambda, N-1$), and is used when performing vector quantization of residual quadrature transformation coefficient $Xresid_k$ 1105 in vector quantization section 1106.

[0216] Also, gain codebook 1109 is composed of previously created N_f kinds of residual gain codes $gainresid^f$ ($f = 0, \Lambda, N_f-1$), and is used when performing vector quantization of residual quadrature transformation coefficient $Xresid_k$ 1105 in vector quantization section 1106.

[0217] The process performed by vector quantization section 1106 will now be described in detail using FIG.12. In step 1201, initialization is performed by assigning

0 to code vector index e in shape codebook 1108, and a sufficiently large value to minimum error $Dist_{MIN}$.

[0218] In step 1202, N-dimensional code vector $coderesid_k^e$ (k = 0, Λ, N-1) is read from shape codebook 1108.

[0219] In step 1203, residual quadrature transformation coefficient $Xresid_k$ output from quadrature transformation processing section 1103 is input, and gain Gainresid of code vector $coderesid_k^e$ (k = 0, Λ, N-1) read in step 1202 is found by means of Equation (51).

[0220]

$$Gainresid = \sum_{k=0}^{N-1} Xresid_k \cdot coderesid_k^e \Big/ \sum_{k=0}^{N-1} {coderesid_k^e}^2 \quad \text{[Equation 51]}$$

[0221] In step 1204, 0 is assigned to $calc\_count_{resid}$ indicating the number of executions of step 1205.

[0222] In step 1205, auditory masking characteristic value M_k output from auditory masking characteristic value calculation section 203 is input, and temporary gain $temp2_k$ (k = 0, Λ, N-1) is found by means of Equation (52).

[0223]

$$temp2_k = \begin{cases} coderesid_k^e & \left(\left|coderesid_k^e \cdot Gainresid + Xbase_k\right| \geq M_k\right) \\ 0 & \left(\left|coderesid_k^e \cdot Gainresid + Xbase_k\right| < M_k\right) \end{cases} \quad (k = 0, \cdots, N-1)$$

[Equation 52]

[0224] In Equation (52), if k satisfies the condition $|coderesid_k^e \cdot Gainresid + Xbase_k| \geq M_k$, $coderesid_k^e$ is assigned to temporary gain $temp2_k$, and if k satisfies the condition $|coderesid_k^e \cdot Gainresid + Xbase_k| < M_k$, 0 is

assigned to $temp2_k$. Here, k is the index of each sample in one frame.

[0225] Then, in step 1205, gain Gainresid is found by means of Equation (53).

[0226]

$$Gainresid = \sum_{k=0}^{N-1} Xresid_k \cdot temp2_k \Big/ \sum_{k=0}^{N-1} temp2_k^{\ 2} \quad (k = 0, \cdots, N-1) \text{[Equation 53]}$$

[0227] If temporary gain $temp2_k$ is 0 for all k's, 0 is assigned to gain Gainresid. Also, residual coded value $Rresid_k$ is found from gain Gainresid and code vector $coderesid_k^e$ by means of Equation (54).

[0228]

$$Rresid_k = Gainresid \cdot coderesid_k^e \quad (k = 0, \cdots, N-1) \text{[Equation 54]}$$

[0229] Also, addition coded value $Rplus_k$ is found from residual coded value $Rresid_k$ and base layer quadrature transformation coefficient $Xbase_k$ by means of Equation(55).

[0230]

$$Rplus_k = Rresid_k + Xbase_k \quad (k = 0, \cdots, N-1) \text{[Equation 55]}$$

[0231] In step 1206, $calc\_count_{resid}$ is incremented by 1.

[0232] In step 1207, $calc\_count_{resid}$ and a predetermined non-negative integer $Nresid_c$ are compared, and the process flow returns to step 1205 if $calc\_count_{resid}$ is a smaller value than $Nresid_c$, or proceeds to step 1208 if $calc\_count_{resid}$ is greater than or equal to $Nresid_c$.

[0233] In step 1208, 0 is assigned to cumulative error

Distresid, and 0 is also assigned to sample index k. Also, in step 1208, addition MDCT coefficient $Xplus_k$ is found by means of Equation (56).

[0234]

$$Xplus_k = Xbase_k + Xresid_k \qquad (k = 0, \cdots, N-1) \text{[Equation 56]}$$

[0235] Next, in steps 1209, 1211, 1212, and 1214, case determination is performed for the relative positional relationship between auditory masking characteristic value M_k 1107, addition coded value $Rplus_k$, and addition MDCT coefficient $Xplus_k$, and distance calculation is performed in step 1210, 1213, 1215, or 1216 according to the case determination result. This case determination according to the relative positional relationship is shown in FIG.13. In FIG.13, a white circle symbol (○) signifies an addition MDCT coefficient $Xplus_k$, and a black circle symbol (●) signifies an addition coded value $Rplus_k$. The concepts in FIG.13 are the same as explained for FIG.6 in Embodiment 1.

[0236] In step 1209, whether or not the relative positional relationship between auditory masking characteristic value M_k, addition coded value $Rplus_k$, and addition MDCT coefficient $Xplus_k$ corresponds to "Case 1" in FIG.13 is determined by means of the conditional expression in Equation (57).

[0237]

$$\left(|Xplus_k| \geq M_k\right) \quad and \quad \left(|Rplus_k| \geq M_k\right) \quad and \quad \left(Xplus_k \cdot Rplus_k \geq 0\right)$$

[Equation 57]

[0238] Equation (57) signifies a case in which the absolute value of addition MDCT coefficient $Xplus_k$ and the absolute value of addition coded value $Rplus_k$ are both greater than or equal to auditory masking characteristic value M_k, and addition MDCT coefficient $Xplus_k$ and addition coded value $Rplus_k$ are the same codes. If auditory masking characteristic value M_k, addition MDCT coefficient $Xplus_k$, and addition coded value $Rplus_k$ satisfy the conditional expression in Equation (57), the process flow proceeds to step 1210, and if they do not satisfy the conditional expression in Equation (57), the process flow proceeds to step 1211.

[0239] In step 1210, error $Distresid_1$ between $Rplus_k$ and addition MDCT coefficient $Xplus_k$ is found by means of Equation (58), error $Distresid_1$ is added to cumulative error Distresid, and the process flow proceeds to step 1217.

[0240]

$$\begin{aligned} Distresid_1 &= Dresid_{11} \\ &= \left|Xresid_k - Rresid_k\right| \end{aligned} \quad \text{[Equation 58]}$$

[0241] In step 1211, whether or not the relative positional relationship between auditory masking characteristic value M_k, addition coded value $Rplus_k$, and addition MDCT coefficient $Xplus_k$ corresponds to "Case 5" in FIG.13 is determined by means of the conditional expression in Equation (59).

[0242]

$$\left(\left|Xplus_k\right| < M_k\right) \quad and \quad \left(\left|Rplus_k\right| < M_k\right)$$ [Equation 59]

[0243] Equation (59) signifies a case in which the absolute value of addition MDCT coefficient $Xplus_k$ and the absolute value of addition coded value $Rplus_k$ are both less than auditory masking characteristic value M_k. If auditory masking characteristic value M_k, addition coded value $Rplus_k$, and addition MDCT coefficient $Xplus_k$ satisfy the conditional expression in Equation (59), the error between addition coded value $Rplus_k$ and addition MDCT coefficient $Xplus_k$ is taken to be 0, nothing is added to cumulative error Distresid, and the process flow proceeds to step 1217. If auditory masking characteristic value M_k, addition coded value $Rplus_k$, and addition MDCT coefficient $Xplus_k$ do not satisfy the conditional expression in Equation (59), the process flow proceeds to step 1212.

[0244] In step 1212, whether or not the relative positional relationship between auditory masking characteristic value M_k, addition coded value $Rplus_k$, and addition MDCT coefficient $Xplus_k$ corresponds to "Case 2" in FIG.13 is determined by means of the conditional expression in Equation (60).

[0245]

$$\left(\left|Xplus_k\right| \geq M_k\right) \quad and \quad \left(\left|Rplus_k\right| \geq M_k\right) \quad and \quad \left(Xplus_k \cdot Rplus_k < 0\right)$$

[Equation 60]

[0246] Equation (60) signifies a case in which the

absolute value of addition MDCT coefficient $Xplus_k$ and the absolute value of addition coded value $Rplus_k$ are both greater than or equal to auditory masking characteristic value M_k, and addition MDCT coefficient $Xplus_k$ and addition coded value $Rplus_k$ are different codes. If auditory masking characteristic value M_k, addition MDCT coefficient $Xplus_k$, and addition coded value $Rplus_k$ satisfy the conditional expression in Equation (60), the process flow proceeds to step 1213, and if they do not satisfy the conditional expression in Equation (60), the process flow proceeds to step 1214.

[0247] In step 1213, error $Distresid_2$ between addition coded value $Rplus_k$ and addition MDCT coefficient $Xplus_k$ is found by means of Equation (61), error $Distresid_2$ is added to cumulative error Distresid, and the process flow proceeds to step 1217.

[0248]

$$Distresid_2 = Dresid_{21} + Dresid_{22} + \beta_{resid} * Dresid_{23} \text{ [Equation 61]}$$

[0249] Here, β_{resid} is a value set as appropriate according to addition MDCT coefficient $Xplus_k$, addition coded value $Rplus_k$, and auditory masking characteristic value M_k. A value of 1 or less is suitable for β_{resid}. $Dresid_{21}$, $Dresid_{22}$, and $Dresid_{23}$ are found by means of Equation (62), Equation (63), and Equation (64), respectively.

[0250]

$$Dresid_{21} = |Xplus_k| - M_k \text{ [Equation 62]}$$

[0251]

$$Dresid_{22} = |Rplus_k| - M_k \text{ [Equation 63]}$$

[0252]

$$Dresid_{23} = M_k \cdot 2 \text{ [Equation 64]}$$

[0253] In step 1214, whether or not the relative positional relationship between auditory masking characteristic value M_k, addition coded value $Rplus_k$, and addition MDCT coefficient $Xplus_k$ corresponds to "Case 3" in FIG.13 is determined by means of the conditional expression in Equation (65).

[0254]

$$(|Xplus_k| \geq M_k) \quad and \quad (|Rplus_k| < M_k) \text{ [Equation 65]}$$

[0255] Equation (65) signifies a case in which the absolute value of addition MDCT coefficient $Xplus_k$ is greater than or equal to auditory masking characteristic value M_k, and addition coded value $Rplus_k$ is less than auditory masking characteristic value M_k. If auditory masking characteristic value M_k, addition MDCT coefficient $Xplus_k$, and addition coded value $Rplus_k$ satisfy the conditional expression in Equation (65), the process flow proceeds to step 1215, and if they do not satisfy the conditional expression in Equation (65), the process flow proceeds to step 1216.

[0256] In step 1215, error $Distresid_3$ between addition coded value $Rplus_k$ and addition MDCT coefficient $Xplus_k$ is found by means of Equation (66), error $Distresid_3$ is added to cumulative error Distresid, and the process flow proceeds to step 1217.

[0257]

$$\begin{aligned} Distresid_3 &= Dresid_{31} \\ &= |Xplus_k| - M_k \end{aligned} \quad \text{[Equation 66]}$$

[0258] In step 1216, the relative positional relationship between auditory masking characteristic value M_k, addition coded value $Rplus_k$, and addition MDCT coefficient $Xplus_k$ corresponds to "Case 4" in FIG.13, and the conditional expression in Equation (67) is satisfied.

[0259]

$$(|Xplus_k| < M_k) \quad and \quad (|Rplus_k| \geq M_k) \quad \text{[Equation 67]}$$

[0260] Equation (67) signifies a case in which the absolute value of addition MDCT coefficient $Xplus_k$ is less than auditory masking characteristic value M_k, and addition coded value $Rplus_k$ is greater than or equal to auditory masking characteristic value M_k. In step 1216, error $Distresid_4$ between addition coded value $Rplus_k$ and addition MDCT coefficient $Xplus_k$ is found by means of Equation (68), error $Distresid_4$ is added to cumulative error Distresid, and the process flow proceeds to step 1217.

[0261]

$$\begin{aligned} Distresid_4 &= Dresid_{41} \\ &= |Rplus_k| - M_k \end{aligned} \quad \text{[Equation 68]}$$

[0262] In step 1217, k is incremented by 1.

[0263] In step 1218, N and k are compared, and if k is a smaller value than N, the process flow returns to step 1209. If k is greater than or equal to N, the process

flow proceeds to step 1219.

[0264] In step 1219, cumulative error Distresid and minimum error $Distresid_{MIN}$ are compared, and if cumulative error Distresid is a smaller value than minimum error $Distresid_{MIN}$, the process flow proceeds to step 1220, whereas if cumulative error Distresid is greater than or equal to minimum error $Distresid_{MIN}$, the process flow proceeds to step 1221.

[0265] In step 1220, cumulative error Distresid is assigned to minimum error $Distresid_{MIN}$, e is assigned to $gainresid\_index_{MIN}$, and gain Distresid is assigned to error minimum gain $Distresid_{MIN}$, and the process flow proceeds to step 1221.

[0266] In step 1221, e is incremented by 1.

[0267] In step 1222, total number of vectors N_e and e are compared, and if e is a smaller value than N_e, the process flow returns to step 1202. If e is greater than or equal to N_e, the process flow proceeds to step 1223.

[0268] In step 1223, N_f kinds of residual gain code $gainresid^f$ (f = 0, Λ, N_f-1) are read from gain codebook 1109, and quantization residual gain error $gainresiderr^f$ (f = 0, Λ, N_f-1) is found by means of Equation (69) for all f's.

[0269]

$$gainresiderr^f = \left|Gainresid_{MIN} - gainresid^f\right| \quad \left(f = 0,\cdots,N_f - 1\right)$$ [Equation 69]

[0270] Then, in step 1223, f for which quantization residual gain error $gainresiderr^f$ (f = 0, Λ, N_f-1) is a

minimum is found, and the found f is assigned to $gainresid\_index_{MIN}$.

[0271] In step 1224, $gainresid\_index_{MIN}$ that is the code vector index for which cumulative error Distresid is a minimum, and $gainresid\_index_{MIN}$ found in step 1223, are output to transmission channel 807 as enhancement layer coded information 806, and processing is terminated.

[0272] Next, enhancement layer decoding section 810 will be described using the block diagram in FIG.14. In the same way as shape codebook 1108, shape codebook 1403 is composed of N_e kinds of N-dimensional code vectors $gainresid_k^e$ ($e = 0, \Lambda, N_e-1$, $k = 0, \Lambda, N-1$), and in the same way as gain codebook 1109, gain codebook 1404 is composed of N_f kinds of residual gain codes $gainresid^f$ ($f = 0, \Lambda, N_f-1$).

[0273] Vector decoding section 1401 has enhancement layer coded information 806 transmitted via transmission channel 807 as input, and using $gainresid\_index_{MIN}$ and $gainresid\_index_{MIN}$ as the coded information, reads code vector $coderesid_k^{coderesid\_indexMIN}$ ($k = 0, \Lambda, N-1$) from shape codebook 1403, and also reads code $gainresid^{gainresid\_indexMIN}$ from gain codebook 1404. Then, vector decoding section 1401 multiplies $gainresid^{gainresid\_indexMIN}$ by $coderesid_k^{coderesid\_indexMIN}$ ($k = 0, \Lambda, N-1$), and outputs $gainresid^{gainresid\_indexMIN} \cdot coderesid_k^{coderesid\_indexMIN}$ ($k = 0, \Lambda, N-1$) obtained as a result of the multiplication to a residual quadrature transformation processing section

1402 as a decoded residual quadrature transformation coefficient.

[0274] The process performed by residual quadrature transformation processing section 1402 will now be described.

[0275] Residual quadrature transformation processing section 1402 has an internal buffer $bufresid_k'$, and initializes this buffer in accordance with Equation (70).

[0276]

$$bufresid'_k = 0 \quad (k = 0, \cdots, N-1) \text{[Equation 70]}$$

[0277] Decoded residual quadrature transformation coefficient $gainresid^{gainresid\_indexMIN} \cdot coderesid_k^{coderesid\_indexMIN}$ (k = 0, Λ, N-1) output from vector decoding section 1401 is input, and enhancement layer decoded signal $yresid_n$ 811 is found by means of Equation (71).

[0278]

$$yresid_n = \frac{2}{N} \sum_{k=0}^{2N-1} Xresid'_k \cos\left[\frac{(2n+1+N)(2k+1)\pi}{4N}\right] \quad (n = 0, \cdots, N-1)$$

[Equation 71]

[0279] Here, $Xresid_k'$ is a vector linking decoded residual quadrature transformation coefficient $gainresid^{gainresid\_indexMIN} \cdot coderesid_k^{coderesid\_indexMIN}$ (k = 0, Λ, N-1) and buffer $bufresid_k'$, and is found by means of Equation (72).

[0280]

$$Xresid'_k = \begin{cases} bufresid'_k & (k = 0, \cdots N-1) \\ gainresid^{gainresid\_index_{MIN}} \cdot coderesid_{k-N}^{coderesid\_index_{MIN}} & (k = N, \cdots 2N-1) \end{cases}$$

[Equation 72]

[0281] Buffer $bufresid_k'$ is then updated by means of Equation (73).

[0282]

$$bufresid'_k = gainresid^{gainresid\_index_{MIN}} \cdot coderesid_k^{coderesid\_index_{MIN}} \quad (k = 0, \cdots N-1)$$

[Equation 73]

[0283] Enhancement layer decoded signal $yresid_n$ 811 is then output.

[0284] The present invention has no restrictions concerning scalable coding layers, and can also be applied to a case in which vector quantization using an auditory masking characteristic value is performed in an upper layer in a hierarchical voice coding and decoding method with three or more layers.

[0285] In vector quantization section 1106, quantization may be performed by appllying acoustic weighting filters to distance calculations in above-described Case 1 through Case 5.

[0286] In this embodiment, a CELP type voice coding and decoding method has been described as the voice coding and decoding method of the base layer coding section and decoding section by way of example, but another voice coding and decoding method may also be used.

[0287] Also, in this embodiment, an example has been given in which base layer coded information and

enhancement layer coded information are transmitted separately, but a configuration may also be taken, whereby coded information of each layer is transmitted multiplexed, and demultiplexing is performed on the receiving side to decode the coded information of each layer.

[0288] Thus, in a scalable coding system, also, applying vector quantization that uses an auditory masking characteristic value of the present invention makes it possible to select a suitable code vector that minimizes degradation of a signal that has a large auditory effect, and obtain a high-quality output signal.

[0289] (Embodiment 3)

FIG.15 is a block diagram showing the configuration of a voice signal transmitting apparatus and voice signal receiving apparatus containing the coding apparatus and decoding apparatus described in above Embodiments 1 and 2 according to Embodiment 3 of the present invention. More specific applications include mobile phones, car navigation systems, and the like.

[0290] In FIG.15, input apparatus 1502 performs A/D conversion of voice signal 1500 to a digital signal, and outputs this digital signal to voice/musical tone coding apparatus 1503.

Voice/musical tone coding apparatus 1503 is equipped with voice/musical tone coding apparatus 101 shown in FIG.1, codes a digital signal output from input apparatus 1502,

and outputs coded information to RF modulation apparatus 1504. RF modulation apparatus 1504 converts voice coded information output from voice/musical tone coding apparatus 1503 to a signal to be sent on propagation medium such as a radio wave, and outputs the resulting signal to transmitting antenna 1505.

Transmitting antenna 1505 sends the output signal output from RF modulation apparatus 1504 as a radio wave (RF signal). RF signal 1506 in the figure represents a radio wave (RF signal) sent from transmitting antenna 1505. This completes a description of the configuration and operation of a voice signal transmitting apparatus.

[0291] RF signal 1507 is received by receiving antenna 1508, and is output to RF demodulation apparatus 1509. RF signal 1507 in the figure represents a radio wave received by receiving antenna 1508, and as long as there is no signal attenuation or noise superimposition in the propagation path, is exactly the same as RF signal 1506.

[0292] RF demodulation apparatus 1509 demodulates voice coded information from the RF signal output from receiving antenna 1508, and outputs the result to voice/musical tone decoding apparatus 1510. Voice/musical tone decoding apparatus 1510 is equipped with voice/musical tone decoding apparatus 105 shown in FIG.1, and decodes a voice signal from voice coded information output from RF demodulation apparatus 1509. Output apparatus 1511 performs D/A conversion of the decoded digital voice

signal to an analog signal, converts the electrical signal to vibrations of the air, and outputs sound waves audible to the human ear.

[0293] Thus, a high-quality output signal can be obtained in both a voice signal transmitting apparatus and a voice signal receiving apparatus.

[0294] The present application is based on Japanese Patent Application No.2003-433160 filed on December 26, 2003, the entire content of which is expressly incorporated herein by reference.

Industrial Applicability

[0295] The present invention has advantages of selecting a suitable code vector that minimizes degradation of a signal that has a large auditory effect, and obtaining a high-quality output signal by applying vector quantization that uses an auditory masking characteristic value. Also, the present invention is applicable to the fields of packet communication systems typified by Internet communications, and mobile communication systems such as mobile phone and car navigation systems.